提高公共数据利用与科研人员数据意识能力的双重挑战

顾立平 [12]
1.中国科学院文献情报中心, 北京, 100190
2.中国科学院大学经济管理学院信息资源管理系, 北京, 100190

【摘要】 本文探索公共数据管理的分层管理以及全球各地相关机构制定的数据开放共享政策。科学数据在数字化科研时代已成为重要战略资源，其开放和共享对于推动科技进步和创新具有关键价值。各国政府、科研机构、学术出版界、信息服务机构等学术社群成员需要共同努力，推动公共数据的开放利用，以实现数据的价值和效用最大化。科研教育机构在数据管理中扮演着双重角色，需要探索有效的数据管理方式，确保数据的可发现性、可访问性、可互操作性和长期保存，同时平衡数据的开放利用与安全保密。为实现这些目标，机构需广泛征求科研团队意见，建立科学数据管理运行的软硬件设施、资金和人员保障，并制定数据管理和共享的标准规范，提供数据管理和共享的培训和支持，提高科研人员的数据意识和能力。

【关键词】科学数据管理；数据开放共享；公共数据；科研教育机构；数据意识能力

Dual challenge of improving public data utilization and data awareness among researchers

GU Liping[12]
1.National Science Library, Chinese Academy of Sciences
2.Department of Information Rescource Management, School of Economic and Management, University of Chinese Academy of Sciences

【Abstract】: This article discusses the hierarchical management of public data management and the open data sharing policies formulated by relevant institutions around the world. Scientific data has become an important strategic resource in the era of digital research, and its openness and sharing have critical value in promoting technological progress and innovation. Governments, research institutions, academic publishing communities, information service institutions, and other members of the academic community need to work together to promote the open utilization of public data, in order to maximize the value and utility of data. Research and education institutions play a dual role in data management, and need to explore effective data management methods to ensure the discoverability, accessibility, interoperability, and long-term preservation of data, while balancing open utilization and security of data. To achieve these goals, institutions need to widely solicit opinions from research teams, establish software and hardware facilities, funding, and personnel support for scientific data management and operation, and develop standard specifications for data management and sharing, provide training and support for data management and sharing, and improve the data awareness and ability of research personnel.

【Keywords】Scientific data management, open data sharing, public data, research and education institutions, data awareness capabilities

一、公共数据管理的分层管理

科学数据在数字化科研时代已成为重要的战略资源，其开放和共享对于推动科技进步和创新具有关键价值。完善的数据共享政策不仅为科学数据管理提供了良好的指导，更促进了科学数据开放共享的积极实践。全球各地的相关机构已经制定了许多数据开放共享政策，并在实践中不断积累和总结经验。国外相关机构已经制定了许多数据开放共享政策，并且拥有相应的管理实践经验。

科学数据管理工作需要进行国家统筹、各部门与各地区分工负责的体制。2001 年 12 月开放协会研究所（Open Society Institute，简称 OSI）提出了《布达佩斯开放存取计划》（Budapest Open Access Initiative，简称 BOAI）[1]，正式掀起了科学数据开放共享的浪潮。在此基础上，2003 年 6 月《百斯达开放获取出版宣言》（Bethesda Statement on Open Access Publishing）正式公布，重新对开放获取的概念进行了修正[2]。随后，西方发达国家以及部分发展中国家的学术社群纷纷响应，相继制定了开放获取政策。2003 年 10 月德国、法国、意大利等国的科研机构在德国柏林联合签署了由德国马普学会发起的《关于自然科学与人文科学资源的开放获取的柏林宣言》(Open Access to Knowledge in the Sciences and Humanities) [3]，强调学术论文、教育资源、科学数据的开放共享，呼吁各国科研机构向网络使用者免费开放更多科学资源。经济合作与发展组织 (OECD) 于 2006 年颁布了《OECD 关于公共资助的科学数据获取的指导方针和原则》，指出“开放性意味着在同等条件下，国际研究界以最低的成本—最好不超过传播的边际成本—来获取科学数据[4]。公共资助的科学数据的公开获取应该是容易获取的、及时的、用户友好的，并且最好是基于互联网的”。英国皇家学会在 2012 年 5 月发布《科学是开放事业报告》中指出，发布科学理论以及实验和观察数据是其他人判断、同意、拒绝、理解该项工作的基础，政府应制定科学数据开放共享政策，并支持相关软件工具的开发和人员技能的培训工作等[5]。

有此可知，对于公共数据的开放利用，我们需要明确其价值和目标选择。数据开放的目的不仅在于推动科研进步，还在于促进经济社会发展，提高公共服务的效率和质量。因此，我们需要选择合适的开放模式和利用方式，以确保数据的价值和效用得到最大化。

二、国家行政机构的数据政策

国家政府在推进数据共享的进程中扮演着至关重要的角色。政府数据的开放共享不仅有助于提高政府透明度，还能促进经济社会发展，提升公共服务的质量和效率。因此，包括我国在内的主要科技大国政府纷纷制定相关政策并展开工作，政府数据开放共享的发展环境也日渐成熟。

美国白宫科技政策办公室（OSTP）2013 年 2 月的开放政府政策指出：“无论全部或者部分受到联邦资助的科研项目，所产生的数字形式的科学数据都应该存储起来，并且提供搜索、检索和使用等的公共访问共享” [6]。在此框架下，美国国家科研基金会（NSF）于 2013 年调整了资助政策[7]，并在 2015 年 3 月发布了配合 OSTP 政策的实施细节[8]，又再 2022 年延续和发布这项政策。美国出版商

协会、美国图书馆学会、美国研究信息服务机构学会也制定了各自的“实施方案”，如 SHare 等[9]。此外，美国透过 G8 会议颁布政府开放数据宪章[10]，英国、日本、法国等 G8 成员积极响应，加拿大、印度、新西兰等国政府也分别发布了科学数据开放共享的声明。

美国白宫科技政策办公室在开放政府政策中指出，政府资助的科研项目所产生的数字形式的科学数据应该存储起来，并提供公共访问共享。这一政策体现了公共数据开放利用的价值和目标选择，即促进科学研究的进步和创新，同时提高政府数据的利用效率和社会效益。

在此框架下，美国国家科研基金会调整了资助政策，并发布了配合开放政府政策的实施细节。这些政策不仅要求科研项目产生的数据需要开放共享，还提供了数据管理和共享的标准规范，以及开放共享的评价考核机制。这些措施有助于提高数据的可发现性，促进数据的长期保存和利用，提升了数据资源的利用效率。

此外，美国还通过 G8 会议颁布了政府开放数据宪章，呼吁其他国家共同参与政府数据开放共享的行动。这一行动体现了公共数据开放利用的全球性和协作性，即各国政府需要共同努力，推动公共数据的开放利用，以实现全球科学研究的进步和社会的发展。

科学数据开放共享是数字化科研时代发展的基本要求，为实现建设新型的、开放的学术交流体系，往往国际科学数据中心不仅来自学术界的支持，也需要国家行政机构的带头牵引和支持资助，并且协同科研教育机构、学术出版界、信息服务机构等学术社群成员的联合推动下，提高数据的可发现性，促进数据的长期保存，实施科学数据管理并且提升数据资源的利用效率和保证数据权益。

美国 OSTP 和 NSF 的政策以及后续颁布的各项管理办法显示，在制定数据管理政策的同时，也颁布了标准规范，以及开放共享的评价考核。在已有的数据中心的建设基础上，强化管理和数据中心的发展，并且形成 www.data.gov 等整合性入口网站，统筹推进相关工作的发展。

公共数据的开放利用是实现数字化科研时代发展的基本要求，也是建设新型的、开放的学术交流体系的重要基础。政府、科研机构、学术出版界、信息服务机构等学术社群成员需要共同努力，推动公共数据的开放利用，以实现数据的价值和效用最大化。

三、科研管理机构与资助机构的数据政策

在国家的数据管理以及共享的发展战略以及政策指导下，各国建立的科研管理机构与资助机构致力于推动公共数据的开放利用。这些机构不仅提供研究、咨询与服务的工作，还加强完善有效的激励措施，以促进科研人员共享和利用公共数据。

公共数据的开放利用价值主要体现在促进科学研究进步和创新，提高政府透明度，推动经济社会发展等方面。因此，科研资助机构的数据共享政策成为科研人员和数据持有者行为的基本依据。这些政策旨在保证科学数据的真实性和完整性，同时要求受到资助的科研项目必须长期保存科学数据，以推动数据的共享和利用。

在实践过程中，各国科研资助机构和管理机构加强规范科学数据管理，提出科学数据密级分层的规范，由所属的国际级科学数据中心按照专业领域的特殊性进行划分。这些措施有助于提高数据的可发现性，促进数据的长期保存和利用，提升了数据资源的利用效率。

科研资助机构作为科研项目的主要资助主体，其建议和强制执行会对科研人员以及数据持有方式等产生更大影响[11]。科研资助机构的数据共享政策往往是科研教育机构、科研人员数据行为的基本依据。英国研究理事会（RCUK）在 2005 年 6 月发布了《RCUK 关于研究成果开放获取的立场声明》[12]，2006 年 6 月发布了《RCUK 关于研究成果开放获取的最新立场声明》[13]，2011 年 4 月发布《RUCK 数据政策通用原则》[14]，这些政策对科学数据存储提出了强制性要求，目的在于保证科学数据的真实性和完整性。在此基础上，RCUK 及其 7 个理事会制定了各自的数据管理政策[15] ，普遍要求受到资助的科研项目，其科学数据必须得到长期保存。美国国家科研基金会（NSF）在 2010 年发布了数据管理政策，规定了申请资助的项目申请书必须附带一份不超过两页的“数据管理计划”，该计划应详细描述数据传播与共享的内容[16]。NSF 自 2013 年开始逐步调整资助政策，评估科研人员的学术成果包括但不限于学术论文，也包括数据引用和其他[17]。2015 年 3 月，NSF 发布《Today’s Data, Tomorrow’s Discoveries》呼应 OSTP 的开放数据政策，并且提出数据管理计划要求、数据引用指南、科技报告要求、覆盖出版和筹备数据存储与管理经费的政策等[18]。此外，英国惠康基金会（WT）[19]、英国癌症研究基金（CRUK）[20]、美国国家卫生研究院（NIH）[21]等也制定了相关的数据管理、保存和共享政策，为开放获取事业推进创造了良好的环境。

这些机构还积极探索有效的激励措施，鼓励科研人员共享和利用公共数据。例如，一些机构将数据共享作为科研项目申请和评估的重要指标，为科研人员提供数据管理和共享的培训和支持等。

公共数据的开放利用是实现数字化科研时代发展的基本要求，也是推动科学进步和创新的重要基础。各国科研资助机构和管理机构应继续加强合作，完善政策规范，提高数据质量，推动公共数据的开放利用，以实现数据的价值和效用最大化。

四、科研教育机构的数据管理政策

在科研管理机构与科研资助机构的数据政策的指导之下，科研教育机构涉及到具体实施单位的数据管理能力，因此科研教育机构在数据管理中扮演着双重角色，一方面是科研资助机构资助的主要对象，需要遵从科研资助机构的数据政策要求，另一方面也要对其科研人员创造的科学数据进行规范管理。

在科研管理机构与科研资助机构的数据政策的指导之下，科研教育机构扮演着重要的角色，不仅需要遵从科研资助机构的数据政策要求，还要对其科研人员创造的科学数据进行规范管理。因此，科研教育机构需要制定自己的管理规范，以确保数据的质量和安全。

公共数据的开放利用价值在于促进科学研究进步和创新，提高政府透明度，推动经济社会发展等。因此，科研教育机构需要积极探索有效的数据管理方式，以确保数据的可发现性、可访问性、可互操作性和长期保存。同时，科研教育机构还需要考虑如何平衡数据的开放利用与数据的安全保密之间的关系，确保数据不会被滥用或泄露。

为了实现这些目标，科研教育机构需要广泛征求科研团队的意见，建立科学数据管理运行的软硬件设施、资金和人员保障。例如，可以建立机构知识库、数据知识库或开放知识库等，提供数据存储、共享和传播的平台。同时，还可以制定数

据管理和共享的标准规范，提供数据管理和共享的培训和支持，提高科研人员的数据意识和能力。

以英国牛津、曼彻斯特、国王、华威、格林多、东伦敦等 7 所科研教育机构为例，在其科学数据政策中，明确表示遵从英国研究理事会数据政策共同原则（2011）（the RCUK Common Principles on Data Policy (2011)）[22]，也分别对内部科学数据的存储、质量保障、传播等进行了规范。如牛津声明科研项目管理中安全存储科学数据、定期备份科学数据至关重要，并为牛津的工作人员、高级科研人员和研究生提供科学数据备份服务[23]；爱丁堡建议科学数据无论其存在的形式是什么，都要至少保持保留 3 份备份，即原始数据版本、外部/本地存储版本、外部/远程存储版本[24]；曼彻斯特声明对机密性数据赋予相关科研人员及其合作方有限时间段内的数据访问特权[25]等。美国哈佛大学信息技术学院（HUIT）与哈佛大学图书馆、定量社会科学研究所（IQSS）、史密森天体物理中心（CfA）合作创建了哈佛大学 Dataverse 网络平台，其数据政策包括用户条款、数据存储条款、数据使用条款、数据备份及保存条款等[26]。

科研教育机构在公共数据开放利用中扮演着重要的角色，需要积极探索有效的数据管理方式，提高数据质量和安全，推动科学研究的进步和创新。

科研教育机构在制定这些条款的同时，也广泛争取科研团队的意见，在条件许可的情况下，进行科学数据管理运行所需软硬件设施等条件、资金和人员保障。例如，建立机构知识库、数据知识库，或者开放知识库等，或者建立指南和推荐存储的列表等。

五、数据知识库的数据政策

数据知识库作为存储科学数据的重要场所，不仅具备了承担相关领域科学数据的整合汇交的基本功能，还在扩展过程中逐步进行科学数据的分级分类、加工整理和分析挖掘等工作。这些数据知识库的建立和发展，为公共数据的开放利用提供了重要的支撑和保障。

数据知识库是具体存储科学数据的位置，多个数据知识库可以形成一个较大范围的数据中心，并且通过联盟形式，扩展为国际科学数据中心。在许多领域的国际科学数据中心，都是逐步从数据知识库发展起来的。

因此，数据知识库本身，就具备了承担相关领域科学数据的整合汇交的基本功能，在扩展的过程当中，逐步进行科学数据的分级分类、加工整理和分析挖掘；等工作。最后通过国内外科学数据方面交流与合作，形成国际科学数据中心。公共数据的开放利用价值主要体现在促进科学研究进步和创新，提高政府透明度，推动经济社会发展等方面。因此，数据知识库的政策制定需要充分考虑这些价值和目标，以促进公共数据的共享和利用。

在具体政策制定方面，数据知识库需要符合所属司法管辖区的各国法律法规，同时也要与科研生态体系相结合。例如，数据知识库可以与科研人员签署合作协议，保证一定的存储空间，但超过这个空间便需要收费。这样的政策可以鼓励科研人员共享和利用公共数据，同时也为数据知识库的运营提供一定的资金支持。

由于国家、基金组织、科研机构、期刊等对支撑科研成果的科研数据提出强制性存储要求和鼓励性建议，国际上建设了一批服务于科研社群的高质量数据知识库，如：Edinburgh DataShare[27]、Open Data LMU[28]、PANGAEA[29]、Dryad[30]、figshare[31]等。为了帮助科研人员分辨并遴选合适的数据知识库，进

而存储和重用其科研数据，数据知识库注册与目录系统应运而生，主要有OAD[32]、re3data.org[33]等，它们以全新的方式对数据知识库的相关信息进行整合、链接和呈现[34]。

数据知识库的政策，在一方面要符合所属司法管辖区的各国法律法规，也要能够与科研生态体系相结合。例如 Dryad 对其存储的数据内容与元数据全部实施 CC0 许可协议，对于数据集则是与数据提交者签署合作协议，并且采取一定的收费形式。无论是 Dryad 还是 figshare 等数据知识库，均保证了一定的存储空间，但超过这个空间便要收费。另一方面，数据存储涉及不同版本的数据集，因此政策中，往往规定作者可以修改元数据，但更改或者增减数据，则需要通过申请或者额外付费等。

数据知识库还需要考虑数据的安全和保密问题，制定相应的数据规范和运营模式，确保数据不会被滥用或泄露。同时，数据知识库还需要具备系统安全保障，确保数据的稳定性和可靠性。数据知识库的政策制定需要充分考虑公共数据的开放利用价值和目标选择，以促进科学研究的进步和创新，提高数据的利用效率和保证数据权益。

数据知识库的数据政策是基础也是核心，它能进行一系列对于未来发展为国际科学数据中心的前景，提供最为关键的元数据、数据集、数据内容、数据规范、运营模式、系统安全保障的原型。在这一方面，已有许多国际规范和实际案例，如上所示，可以借鉴参考，并且与现有的国际数据中心的历史发展进行比较。

六、结语

公共数据的开放利用价值在于推动科技进步和创新，提高政府透明度，促进经济社会发展等。因此，科学数据管理涉及不同层级的管理和服务机构，包括国家、科研管理机构、资助机构、科研教育机构和数据知识库等，它们共同致力于推动公共数据的开放利用。

各国政府在推进数据共享的进程中扮演着关键角色，制定相关政策并展开工作，以促进数据的共享和利用。科研资助机构作为科研项目的主要资助主体，其建议和强制执行会对科研人员以及数据持有方式等产生重要影响，因此需要引导科研人员共享和利用公共数据。

科研教育机构在数据管理中扮演着双重角色，一方面需要遵从科研资助机构的数据政策要求，另一方面也要对其科研人员创造的科学数据进行规范管理。数据知识库是具体存储科学数据的位置，其政策需符合所属司法管辖区的各国法律法规，并与科研生态体系相结合，以确保数据的可发现性、可访问性、可互操作性和长期保存。

总之，各层级管理和服务机构需要共同努力，探索有效的数据管理方式，推动公共数据的开放利用，以实现数据的价值和效用最大化。

参考文献

[01] Budapest Open Access Initiative [OL].[2023-09-22] http://www.budapestopenaccessinitiative.org/
[02] Bethesda Statement on Open Access Publishing [OL].[2023-09-22] http://dash.harvard.edu/bitstream/handle/1/4725199/suber_bethesda.htm?sequence=1

[03] Berlin Declaration. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [OL].[2023-09-22] http://oa.mpg.de/openaccess-berlin/berlindeclaration.html.
[04] Organisation for Economic Co-operation and Development (OECD) .Principles and Guidelines for Access to Research Data from Public Funding [EB/OL]. [2023-09-22] http://www.oecd.org/dataoecd/9/61/385008 13.pdf.
[05] The Royal Society. Science as an open enterprise [EL/OL]. [2023-09-22].http://royalsociety.org/policy/projects/science-public-enterprise/report/.
[06] Office of Science and Technology Policy. Increasing Access to the Results of Federally Funded Scientific Research [EB/OL]. [2023-09-22].https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf#:~:text=Since%20February%202013%2C%20federal%20public%20access%20policy%20has,publications%20and%20digital%20data%20resulting%20from%20such%20research.
[07] NSF. Dissemination and sharing of research results [EB/OL]. [2023-09-22]. http://www.nsf.gov/bfa/dias/policy/dmp.jsp.
[08] NSF. Today’s data, tomorrow’s discoveries [EB/OL]. [2023-09-22]. http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf.
[09] Covey T.D. Publishers and universities respond to the OSTP mandate [EB/OL]. [2013-07-10]. http://works.bepress.com/denise_troll_covey/72.
[10] Cabinet, Office. G8 Open Data Charter and Technical Annex [EB/OL]. [2015-03-29]. https://www.gov.uk/government/publications/open-data-charter/g8-.
[11] The Royal Society. Science as an open enterprise [OL]. [2023-09-22]. .
[12] RCUK. Rcuk position statement on access to research outputs [EB/OL]. [2023-09-22]. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/2005statement.pdf
[13] RCUK. Research councils UK’ updated position statement on access to research outputs [EB/OL]. [2023-09-22]. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/2006statement.pdf
[14] RCUK. Rcuk common principles on data policy [EB/OL]. [2023-09-22]. http://www.rcuk.ac.uk/research/datapolicy/
[15] DCC. Funders' data policies [EB/OL]. [2023-09-22]. http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies.
[16] NSF. Dissemination and sharing of research results [EB/OL]. [2023-09-22].
[17] NSF. Data Management Plan [OL]. [2023-09-22]. http://nsf.gov/pubs/2013/nsf13546/nsf13546.htm.
[18] NSF. Today’s data, Tomorrow’s Discoveries [OL].[2023-09-22] http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf
[19] WT. Policy on data management and sharing [EB/OL]. [2023-09-22].http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm
[20] CRUK. Data sharing policy [EB/OL]. [2023-09-22]. http://www.cancerresearchuk.org/sites/default/files/data_sharing_policy.pdf.

[21] NIH. NIH Data sharing policy and implementation guidance [EB/OL]. [2023-09-22].https://sharing.nih.gov/data-management-and-sharing-policy
[22] RCUK. Data management and sharing policy [EB/OL]. [2023-09-22]. http://www.rcuk.ac.uk/research/datapolicy/.
[23] The University of Oxford. Backup, storage & security [EB/OL]. [2023-09-22]. http://researchdata.ox.ac.uk/home/managing-your-data-at-oxford/storage-and-backup.
[24] University of Edinburgh. Research data management guidance- Data storage and backup [EB/OL]. [2023-09-22]. http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/research-data-mgmt/storage-backup.
[25] University of Manchester. Frequently asked questions [EB/OL]. [2023-09-22]. http://www.library.manchester.ac.uk/ourservices/research-services/rdm/faq/.
[26] IQSS. Dataverse.[OL].[2023-09-22] http://dvn.iq.harvard.edu
[27] Edinburgh DataShare[EB/OL].[2023-09-22]. http://datashare.is.ed.ac.uk/.
[28] University of Muenchen. Open Data LMU[EB/OL].[2023-09-22]. http://data.ub.uni-muenchen.de/
[29] PANGAEA. About [EB/OL].[2023-09-22]. https://www.pangaea.de/about/
[30] Dryad. Terms of Service. [EB/OL].[2023-09-22]. http://datadryad.org/pages/policies.
[31] Figshare. About [EB/OL].[2023-09-22]. https://figshare.com/about.
[32] OAD. Editorial Policy. [EB/OL].[2023-09-22]. http://oad.simmons.edu/oadwiki/Editorial_Policy
[33] re3data.org[EB/OL].[2023-09-22].http://re3data.org.
[34] IQSS. Dataverse.[OL].[2023-09-22] http://dvn.iq.harvard.edu